

# Modelo determinista para el análisis sintáctico: una perspectiva algorítmica y computacional de la lingüística

Julio Meroño Sáez

December 2, 2023

## Contents

<b>1</b>	<b>Introducción</b>	<b>2</b>
<b>2</b>	<b>Justificación</b>	<b>3</b>
<b>3</b>	<b>Objetivos del trabajo</b>	<b>5</b>
<b>4</b>	<b>Modelo adaptado de la gramática española</b>	<b>6</b>
4.1	Los rasgos lingüísticos . . . . .	6
4.2	Nivel léxico-morfológico: la lexía . . . . .	6
4.2.1	Las lexías semánticas . . . . .	6
4.2.1.1	Adjetivos . . . . .	6
4.2.1.2	Adverbios . . . . .	6
4.2.1.3	Sustantivos . . . . .	6
4.2.1.4	Verbos . . . . .	6
4.2.2	Las lexías determinantes . . . . .	6
4.2.2.1	Artículos . . . . .	6
4.2.2.2	Cuantificadores . . . . .	6
4.2.2.3	Demostrativos . . . . .	6
4.2.2.4	Interrogativos . . . . .	6
4.2.2.5	Numerales . . . . .	6
4.2.2.6	Posesivos . . . . .	6
4.2.2.7	Pronombres personales . . . . .	6
4.2.2.8	Relativos . . . . .	6
4.2.3	Las lexías gramaticales . . . . .	6
4.2.3.1	Conjunciones . . . . .	6

4.2.3.2	Preposiciones . . . . .	6
4.3	Nivel sintáctico interno: el sintagma . . . . .	7
4.3.1	El sintagma adjetival . . . . .	7
4.3.2	El sintagma adverbial . . . . .	7
4.3.3	El sintagma nominal . . . . .	7
4.3.4	El sintagma verbal . . . . .	7
4.4	Nivel sintáctico externo: la cláusula . . . . .	7
<b>5</b>	<b>Implementación del modelo</b>	<b>7</b>
<b>6</b>	<b>Accesibilidad al público</b>	<b>7</b>
<b>7</b>	<b>Evaluación del modelo</b>	<b>7</b>
<b>8</b>	<b>Conclusiones</b>	<b>7</b>

# 1 Introducción

El español o castellano es una lengua iberorromance y es la lengua nativa de más de 450 millones de hablantes en el mundo, a lo que se le suman los 75 millones de hablantes no nativos alrededor del planeta. Esto la sitúa como la cuarta lengua más hablada globalmente, detrás del inglés, el chino mandarín y el hindi. Es por ello que la existencia de una lingüística estandarizada es fundamental, al permitir la comunicación internacional sin cabida a ningún tipo de ambigüedad y al mantener la integridad de la lengua. La Asociación de Academias de la Lengua Española, junto a sus correspondientes academias nacionales (entre ellas, la Real Academia Española), son las encargadas de dicho cometido, proporcionando una estandarización de la gramática, morfología, léxico y fonética del español alrededor del mundo hispanohablante.

Sin duda, la gramática es tan fundamental como las otras ramas de estudio de la lingüística. Esta es la encargada de estudiar la organización y la función de las palabras dentro de una oración, junto a la agrupación de palabras en unidades mayores con significado y función, denominadas sintagmas. A su vez, dichos sintagmas se organizan por categorías sintagmáticas que engloban y marcan su función general. El estudio y el estandarizado de la gramática en el español se remonta a finales del siglo XV, con autores como Antonio de Nebrija, llegando y continuando hasta la actualidad con teorías y propuestas transversales a la mayoría de lenguas, como las propuestas por Noam Chomsky, o trabajos más enfocados en el español.

Tradicionalmente, y aún en épocas recientes, se ha realizado un estudio sintáctico más específico para cada categoría sintagmática, donde los sintagmas (o grupos, como se hacen llamar en ciertos modelos) poseen estructuras fundamentalmente distintas entre tipos, y

donde se da lugar a numerosos casos límite a tener en cuenta. Aún así, dicho análisis ha resultado ser considerablemente más efectivo y simple a la hora de su uso superficial por lingüistas procedentes de otras ramas, o en el ámbito educativo como medio de enseñanza. Sin embargo, desde finales del siglo anterior y gracias a la teoría estándar propuesta por Noam Chomsky, se ha dado lugar a otro enfoque distinto a la gramática, denominado gramática generativa. En la mayoría de teorías y modelos asociados con dicha gramática, se formula que todas las agrupaciones a las que se les denomina sintagmas comparten una estructura fundamental, que aun pudiendo variar en el tipo de unidades y tipos de sintagmas agrupados para formarlo, siempre se estructuran de la misma manera, sin casi ninguna cabida a variación.

El estudio de la gramática en el español, especialmente de la sintaxis, ha resultado en el creciente interés por el uso de las nuevas tecnologías y de la capacidad computacional de este siglo para el análisis sintáctico de frases, sintagmas individuales y oraciones. Por desgracia, el uso en su mayoría de la gramática tradicional para la creación de modelos y algoritmos que cumplan dicha tarea dificulta considerablemente la detección y el correcto análisis de ciertos casos especiales o límite que se pueden dar, o imposibilita casi por completo el proporcionamiento de más información en cuanto a la función que realizan ciertos sintagmas dentro de agrupaciones mayores, factor que la gramática tradicional no suele tener en cuenta.

Es por ello por lo que este trabajo de investigación propone un análisis determinista en forma de modelo de la gramática del español, basándose fundamentalmente en la gramática generativista aportada en la obra de Karen Zagana, Sintaxis generativa del español [1], pero incluyendo modificaciones tanto propias como de otros autores que permitan una mayor facilidad para el procesado de oraciones mediante implementaciones computacionales de dicho modelo. Además, se pretende diseñar y llevar a cabo una implementación tipo del modelo formulado previamente que sea capaz de poder utilizarse en ordenadores convencionales sin requerir de un excesivo coste computacional, para su posterior publicación mediante una interfaz en línea y de acceso libre para su uso público.

## 2 Justificación

Son decenas de millones los estudiantes que se encuentran actualmente aprendiendo español en las escuelas. En específico, contando únicamente los estudiantes extranjeros y no procedentes de ningún país hispanohablante, se estima que el número se encuentra en el orden de los 24 millones [2]. Entre toda la materia y contenidos aportados en la clase a dichos estudiantes durante sus horas lectivas, una gran y significativa parte consiste en el estudio de la propia lengua. Además, la utilización del estudio y análisis sintáctico mediante separación en sintagmas o grupos encuadrados se ha vuelto a lo largo de los años el estándar a nivel nacional y, en ciertos casos, internacional. Aunque los estudios que

lo afirmen son relativamente escasos, predominantemente la experiencia ha demostrado a las instituciones que planifican las competencias a tener en cuenta a la hora de enseñar en las aulas que el entendimiento del funcionamiento de la lengua implica un mejor uso de esta, tanto social y oral como formal y escrito, y es fundamental para la preservación del español, razón junto a las anteriores por las que se muestra un interés significativo en su enseñanza en los centros educativos a lo largo de todo el territorio español. Además, se ha de tener en cuenta la reciente y creciente adopción de esta metodología en otros países de habla española, lo que implica una mayor impulsión del conocimiento interno del funcionamiento de la sintaxis española.

Si bien, el uso de la gramática tradicional como herramienta didáctica se remonta a, aproximadamente, el siglo anterior, la cantidad de material de ejercicios para los estudiantes es particularmente escaso, especialmente cuando se pretende comparar con otras materias donde no solo existe mucho más material, sino que es posible su generación procedural y automática. En principio, lo comentado no debería tener ninguna implicación negativa, pues al tratarse de un estudio de la gramática española en forma de sintaxis basada en teorías tradicionales, la generación de ejercicios resulta tan simple como obtener, buscar o producir por la propia cuenta del profesorado oraciones que se adapten al nivel de los estudiantes. Sin embargo, se ha de considerar y tener en cuenta que, para comprobar la correctitud y validez de las respuestas otorgadas por los estudiantes a espera de ser evaluadas, resulta tedioso tener que realizar una comprobación mental lenta y laboriosa de dichos resultados, que puede además dar lugar a numerosos errores de comprobación, no solo beneficiosos sino también perjudiciales en ciertos casos para el alumnado. Por otro lado, desde el punto de vista estudiantil, la comprobación independiente de resultados es completamente imposible, pues se carece de un punto de referencia objetivo para comparar, por lo que estos carecen de una forma de poder practicar de manera efectiva e ilimitada, sin depender del continuo trabajo del profesorado.

Lo que diferencia principalmente a las personas de cualquiera de los dos puntos de vista anteriores con los ordenadores y dispositivos electrónicos es la capacidad de poder tomar decisiones aplicando el entendimiento horizontal, subjetivo y basado en la experiencia y no en reglas a aplicar. Por lo tanto, la tarea de producir análisis sintácticos utilizables en el ámbito de la creación de asistentes basados en la inteligencia artificial se considera un tema popular y continuamente en desarrollo y debate en investigaciones de las últimas décadas.

Este último caso nos lleva al ámbito judicial, donde se muestra una carencia de estandarización intuitiva del significado de los textos legislativos, los contratos y los acuerdos formales, dando lugar a que existan profesiones cuyo único propósito sea el de garantizar la ausencia de vacíos legales a la hora de producir dichos papeles. Todo esto ocurre por una vaga especificación abierta a interpretaciones de la gramática utilizada en este campo.

La motivación de este trabajo de investigación es, entonces, aportar un pequeño pero significativo grano de arena a la base de lo que será el mundo en los años y las décadas por venir, tanto en el ámbito de la educación, como en el de la tecnología, como en el judicial, como en cualquier otro que se pueda plantear (lenguaje matemático, diplomacia, etcétera). Para ello, se plantean diversos objetivos escalonados que se formulan en el apartado a continuación.

### 3 Objetivos del trabajo

El objetivo principal de este trabajo de investigación es, en resumidas cuentas, poder aportar un modelo descriptivo y determinista pero empírico de la gramática del español, basándose en una agrupación de teorías y definiciones de la gramática de autores anteriores y antecedentes, mezclado con un enfoque propio matizado por la mentalidad computacional y matemática, que pueda utilizarse como herramienta de formalización, educación y comprobación, pero además como punto de partida para futuras investigaciones.

Por otro lado, este objetivo ya mencionado da lugar a ciertos enfoques alternativos pero compatibles por los que llevar la investigación, en forma de objetivos secundarios:

1. Crear una herramienta de uso fácil e inmediato para el análisis de oraciones y sintagmas en español, disponible de forma completamente gratuita al público, para su uso por profesorado, estudiantes, políticos, diplomáticos, abogados, etcétera.
2. Aportar una simplificación horizontal basada en rasgos lingüísticos de las estructuras de árbol presentes en la mayoría de corrientes presentes, principalmente del árbol bipartito propuesto por las principales ramas de la gramática generativista, que permita su uso de forma más eficaz a la hora de realizar análisis computacionales de la sintaxis del español, no solo como punto de apoyo para el anterior objetivo sino también como base para modelos de asistencia mediante inteligencia artificial, chatbots y cualquier tecnología similar.
3. Agrupar los esfuerzos provenientes de las distintas corrientes lingüísticas existentes para propulsar la gramática española en un trabajo recopilatorio que, además, introduce ciertas variaciones partiendo de un enfoque más tecnológico, algorítmico y modélico.

## 4 Modelo adaptado de la gramática española

### 4.1 Los rasgos lingüísticos

### 4.2 Nivel léxico-morfológico: la lexía

#### 4.2.1 Las lexías semánticas

##### 4.2.1.1 Adjetivos

##### 4.2.1.2 Adverbios

##### 4.2.1.3 Sustantivos

##### 4.2.1.4 Verbos

#### 4.2.2 Las lexías determinantes

##### 4.2.2.1 Artículos

##### 4.2.2.2 Cuantificadores

##### 4.2.2.3 Demostrativos

##### 4.2.2.4 Interrogativos

##### 4.2.2.5 Numerales

##### 4.2.2.6 Posesivos

##### 4.2.2.7 Pronombres personales

##### 4.2.2.8 Relativos

#### 4.2.3 Las lexías gramaticales

##### 4.2.3.1 Conjunciones

##### 4.2.3.2 Preposiciones

## 4.3 Nivel sintáctico interno: el sintagma

### 4.3.1 El sintagma adjetival

### 4.3.2 El sintagma adverbial

### 4.3.3 El sintagma nominal

### 4.3.4 El sintagma verbal

## 4.4 Nivel sintáctico externo: la cláusula

# 5 Implementación del modelo

# 6 Accesibilidad al público

# 7 Evaluación del modelo

# 8 Conclusiones